..... Fit Hub Gym Center .....

Neeraj Tripathi

20 September 2020

Setting Up a Gym Center in Chicago, Illinois-USA

1. Introduction

1.1 Background

A modern day gymnasium (as gym used to be called way back in Ancient Greece) is a place for indoor physical workout where various equipment and machines are typically used. For some people, a typical gym is a place where you focus on weight lifting and similar activities. Going to a gym has numerous benefits like a great boost in physical and mental health, stress release, confidence in day to day routine work and a place where one can get a change of pace from the normal routine of the day.

In the busy modern life style, health problems are becoming common day by day because of negligence of people towards staying fit. Simple and consistent exercises can keep a lot of health issues at bay. But due to lack of discipline and consistency, and sometimes (though rare) lack of knowledge discourages people from keeping up with a good exercise routine. This behaviour creates the space for organised systems called Gym/ Fitness Centres/ Health Hubs etc. They provide a common place and a perfect environment for workout as they cater to many similar minded people who are conscious about their health. Thus opening gym centres which provide good facilities/services is profitable business consideration for many.

1.2 Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Chicago, Illinois, USA to open a new Gym/Fitness center. Using data science methodology and Machine Learning techniques like clustering, this project aims to provide solutions to answer the business question: What is the best location to open up a new Gym/Fitness center in Chicago?
This project specifically analyses the neighborhoods of Chicago city, Illinois-USA and suggests suitable locations to open up gym centres.

1.3 Stakeholders

This project is useful for business owners/ people who have an interest in opening up a new gym center in Chicago, USA. This is one of the best time to open up fitness centre in this city, as people here are getting increasingly aware towards their health. The health department has also been campaigning their “Healthy Chicago Plan”, so this is the most apt opportunity to grab loyal customers by providing them the quality and type of service they expect from a standard fitness centre.

2. Data

2.1. Source

To approach this business problem, the required data are :

2.1.1. Data about the neighborhoods of Chicago: Name, Zip code, Latitude, Longitude. This data is obtained from two sources:

2.1.1.1. A .csv file containing all the zip codes and corresponding city, coordinates and time zone for each zip code in the State of Illinois, USA. Source: https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/download/?format=csv&refine.state=IL

2.1.1.2. Data containing zip codes and neighborhoods of all zip codes for the city of Chicago. Source: https://www.seechicagorealestate.com/chicago-zip-codes-by-neighborhood.php

2.1.2. Data about the places in each of these neighborhoods (venue data): Restauranctes, shops, hotels, bars, clubs, stores, malls, gyms etc. This data is obtained via Foursquare API.

2.2. Description

The data obtained from source mentioned in 2.1.1.1. contains zip code, state, city, latitude, longitude, time zone and daylight savings time flag as columns for all the zip codes that belong to the State of Illinois, USA. The data obtained from source mentioned in 2.1.1.2. contains zip code and neighborhood names for all the zip codes that belong to the city of Chicago. There are multiple neighborhoods for some zip codes as expected, so only the first neighborhood is extracted after scraping this data from the source in order to preserve a unique zip to neighborhood mapping. The data from both these sources are then merged together to obtain the required data for neighborhoods of Chicago. This dataset has 99 neighborhoods in total.
After the initial preparation, the venue data is obtained for each neighborhood present in the dataset using Foursquare API. This returns the name, coordinates and the category venues for each of the neighborhoods.

2.2.1. Data Preprocessing

The unnecessary features were discarded after the merge step of neighborhood data and the zip code data, and the final dataset thus obtained has the following features: Zip, Neighborhood, Latitude, Longitude. Next, the venue data is obtained. This results in 9900 venues belonging to 250 venue categories, 100 venues for each neighborhood. For the objective of this project, not all categories are important individually, so related categories are clubbed together into some broad category resulting in only 10 categories. Then one hot encoding is applied on the “Category” column to get the dummies indicating which category a venue belongs to. This one hot encoded dataset has 9900 venues mapped to unique zip and neighborhood, which is also later used to obtain the frequency of each type of venue in a neighborhood.